

Arquitectura genómica y polimorfismos genéticos

Susana Balcells

5 febrero 2010

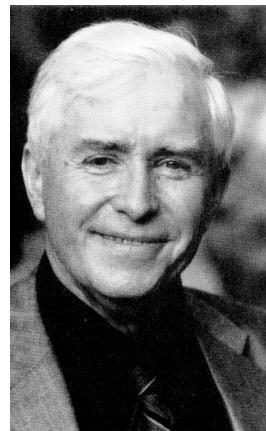
4º curso de genética humana de la SEG

1956: 46 cromosomas humanos



Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics

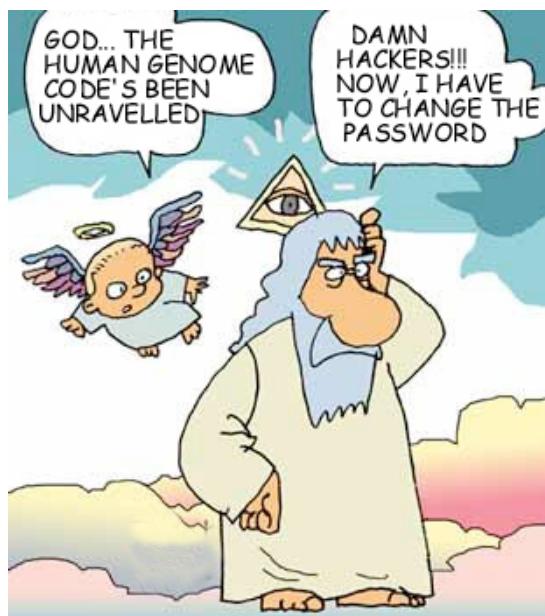
Tjio & Levan. 1956. *Hereditas* 42: 1-6



5 febrero 2010

4º curso de genética humana de la SEG

2000: secuencia del genoma humano



Venter & Collins.

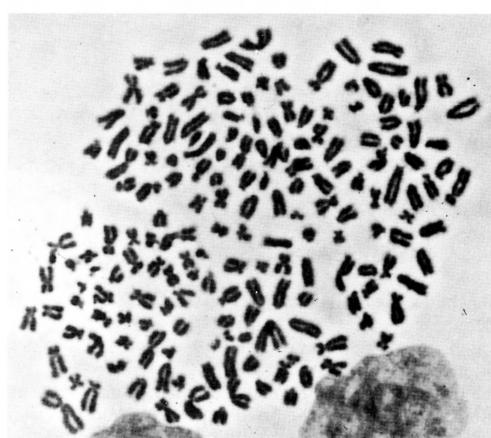
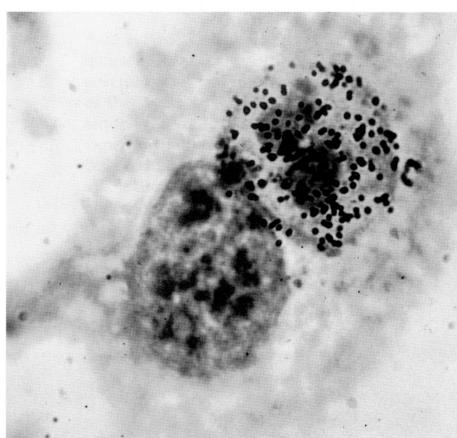


- Hum Genom Consort. 2001. *Nature* 409: 860-921
- Venter et al. 2001. *Science* 291: 1304-1351

5 febrero 2010

4º curso de genética humana de la SEG

1965: Células híbridas humano-ratón

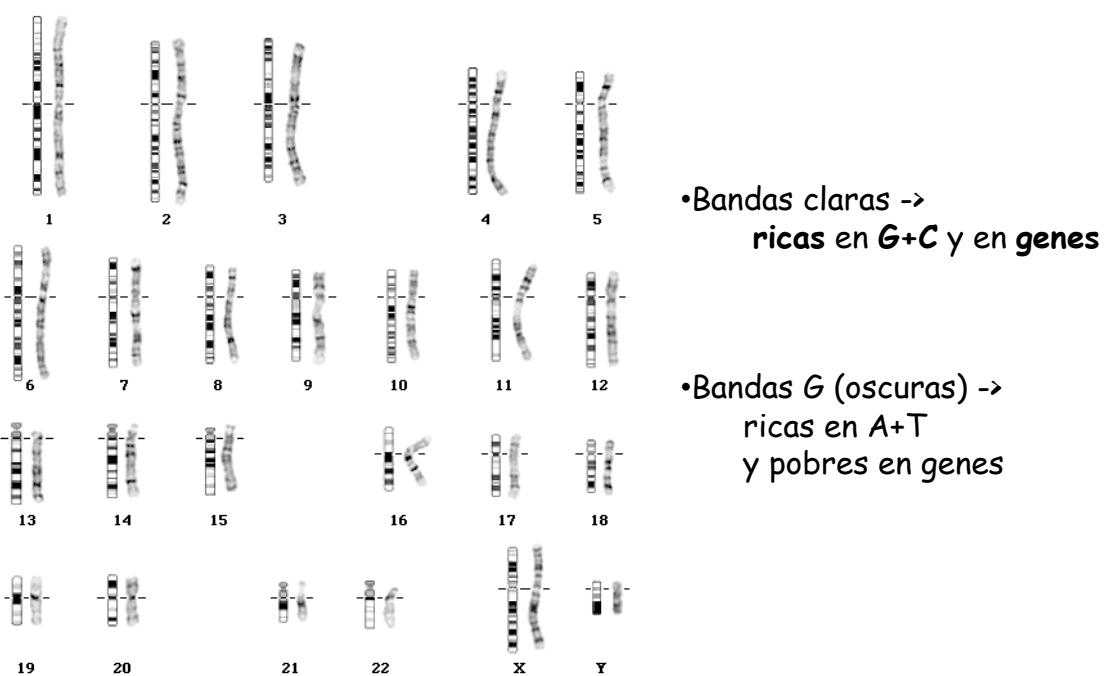


Harris & Watkins. 1965. *Nature* 205: 640-646

5 febrero 2010

4º curso de genética humana de la SEG

≈1971: Bandeo Giemsa



5 febrero 2010

4º curso de genética humana de la SEG

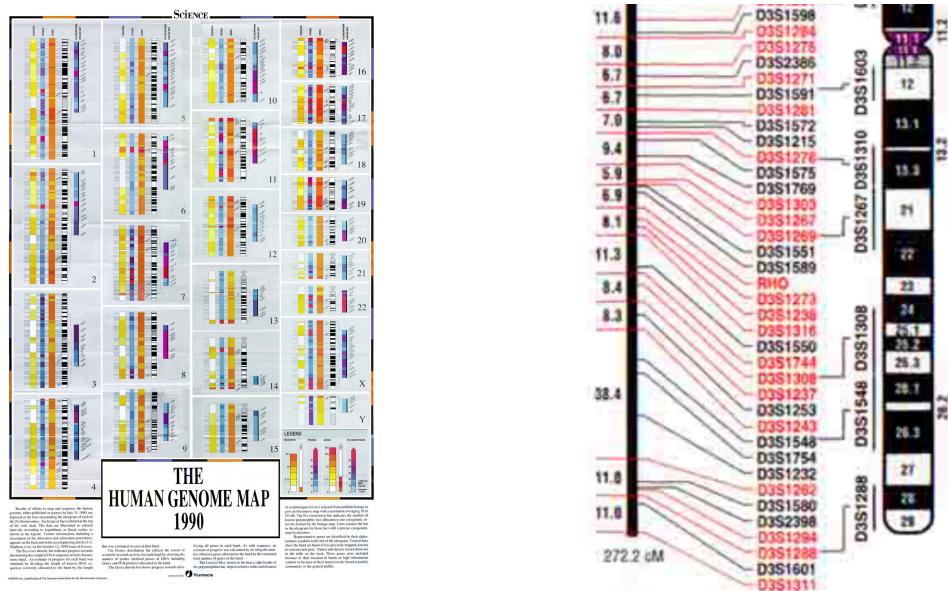


1990: FISH

Fan, Davis & Shows. 1990.
Proc Natl Acad Sci 87:
6223-6227

genética humana de la SEG

1992: mapa genético microsatélites

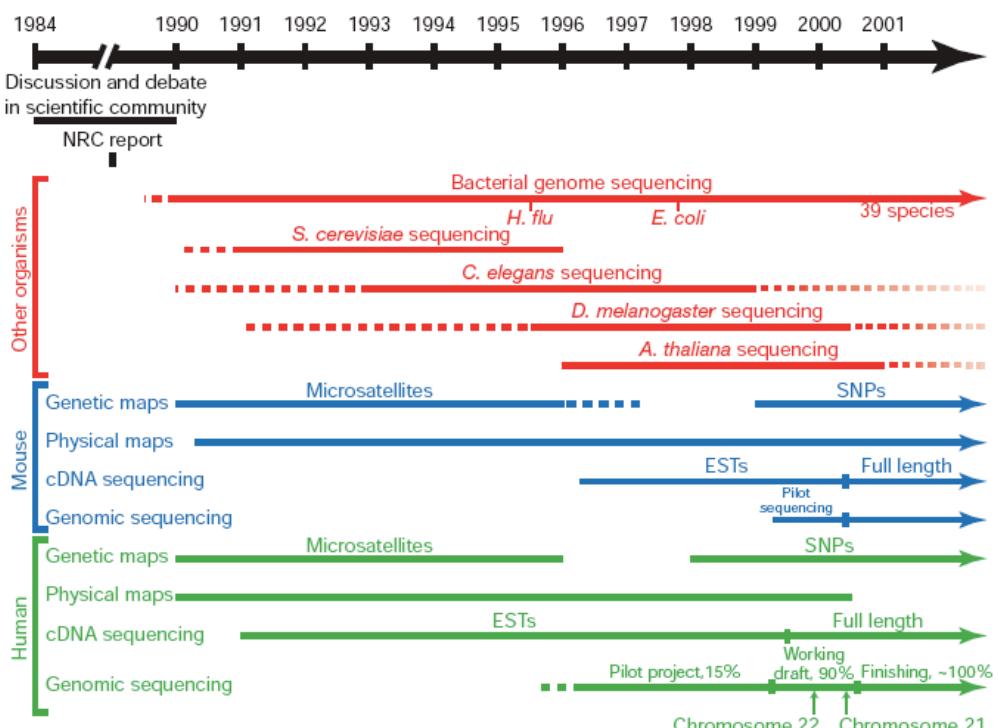


NIH/CEPH Collaborative Mapping Group. 1992. *Science* 258:148-62.

5 febrero 2010

4º curso de genética humana de la SEG

1984-2000: Importantes avances genómicos



5 febrero 2010

4º curso de genética humana de la SEG • Hum Genom Consort. 2001. *Nature* 409: 860-921

3,2 Gigabases = 3200 Megabases

- 3.200.000 kilobases
- 3.200.000.000 pares de bases
- = 700 x genoma *E. coli*
- = 270 x genoma de levadura
- = 19 x genoma de *Drosophila*
- = 1,2 x genoma de ratón

25.000 genes

- Chimpancé 22.524
- Ratón 26.762
- *Drosophila* 13.525
- Levadura 6.144
- *E. Coli* 4.289

Tipos de secuencias

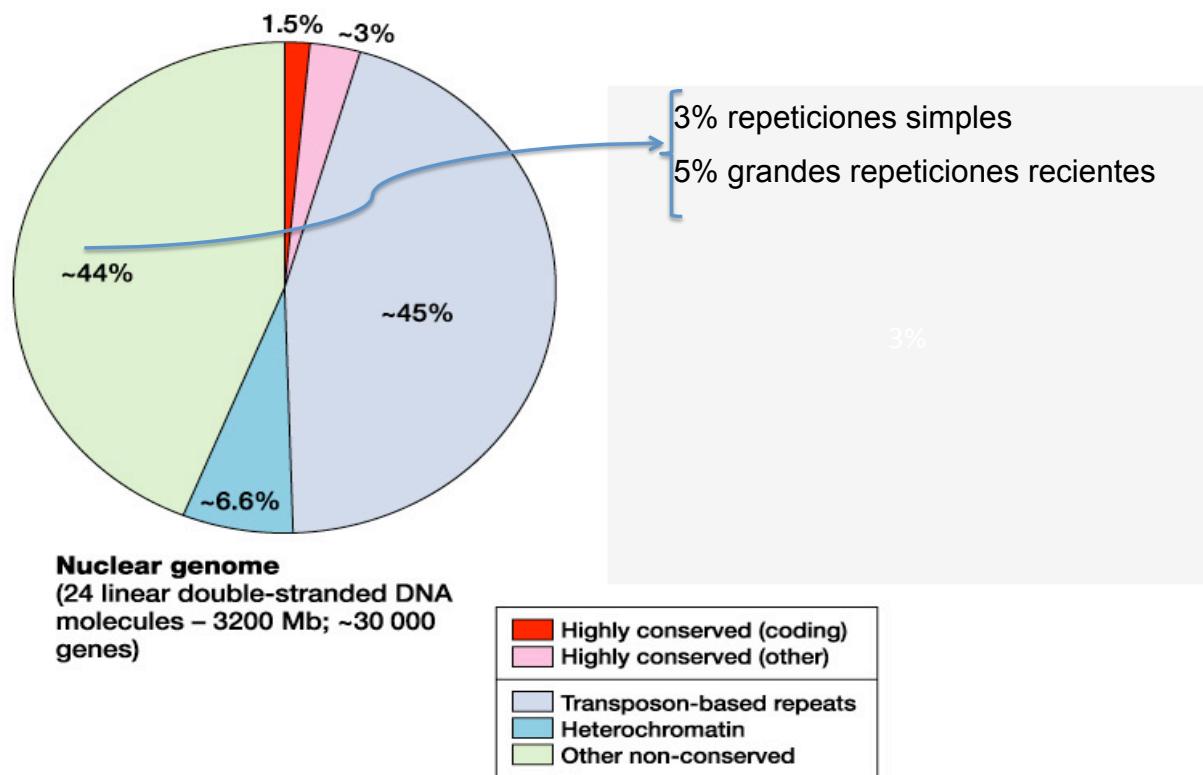
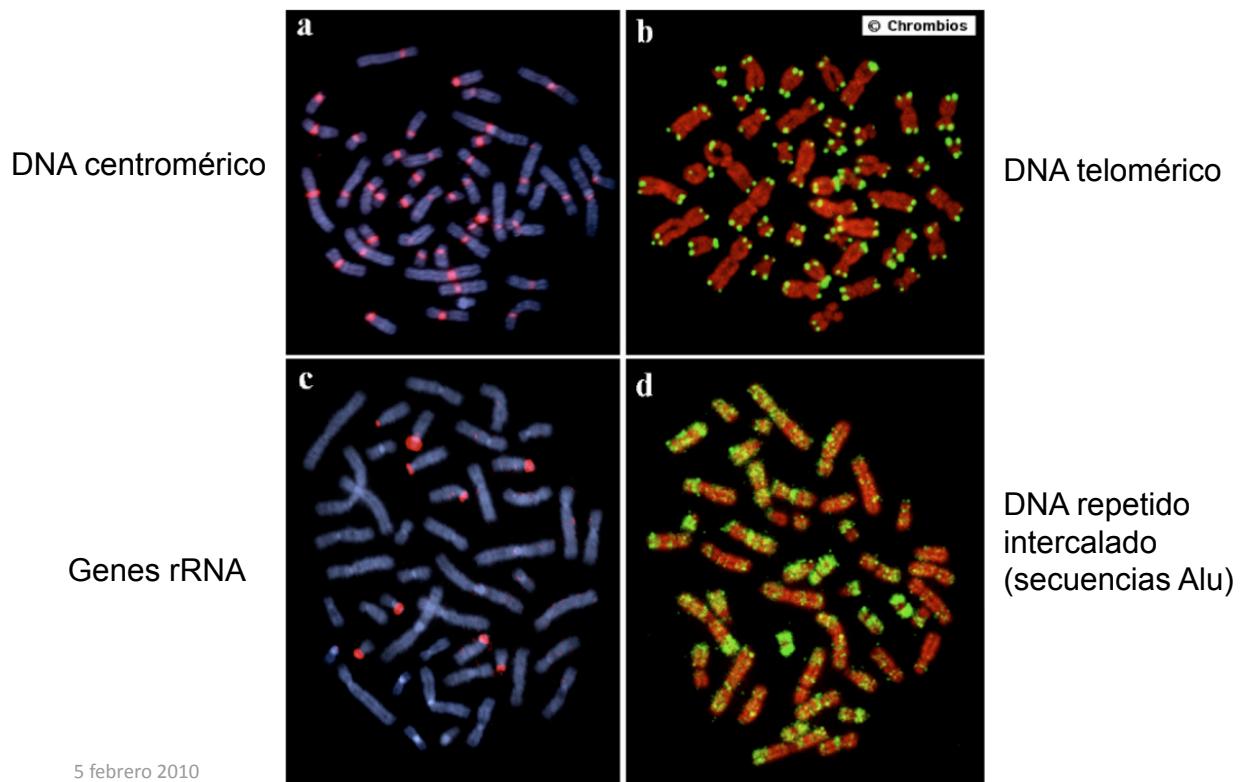


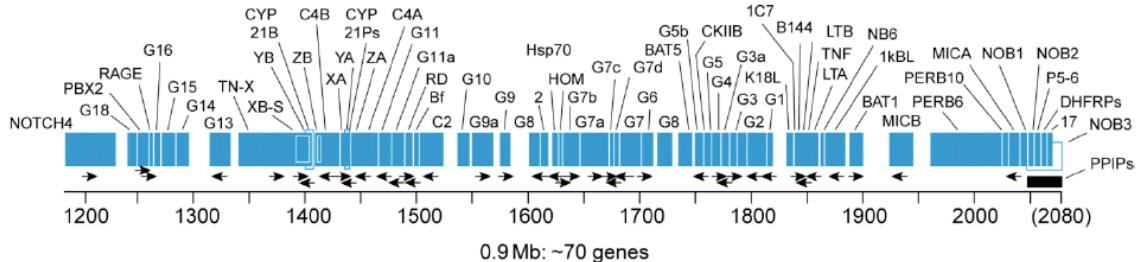
Figure 9-1 Human Molecular Genetics, 3/e. (© Garland Science 2004)

Secuencias repetidas

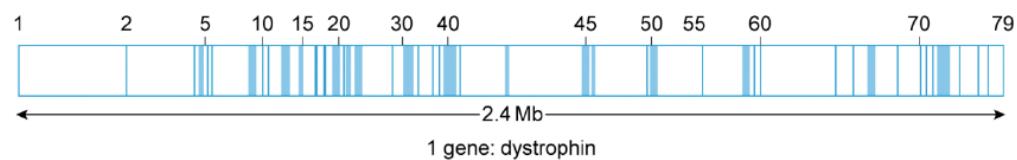


Densidad génica

(A)



(B)



5 febrero 2010

4º curso de genética humana de la SEG

“Polimorfismo de DNA”

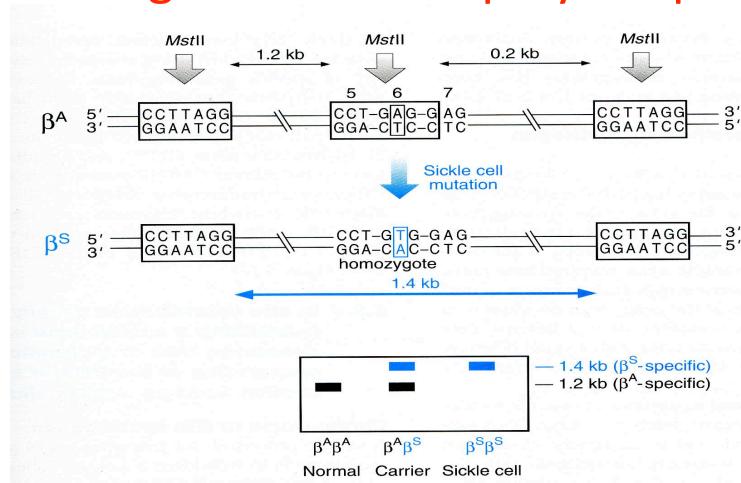
- Locus del genoma humano con más de un alelo
- Con una frecuencia del alelo minoritario de 1% o más en la población

Clases de polimorfismos

- SNP
- VNTR simple
- Polimorfismo de copia de transposición
- VNTR a gran-escala
- Polimorfismo de Inversión
- Polimorfismos cromosómicos

} Variación estructural

SNP: single nucleotide polymorphism



- sustitución, inserción o delección de 1 nt
- \Rightarrow 2 alelos
- Pueden generar RFLPs
- Distribución por todo el genoma
- Genotipado automático a gran escala

SNPs

- Bialélicos
- Frecuencia alelo minoritario (MAF) > 1%
- 10 millones SNPs en el Genoma Humano (1/300 bp)
- Constituyen 90% de la variación en la población (10% restante: variantes minoritarias *rare variants*)

Proyecto HapMap



International HapMap Project

[Home](#) | [About the Project](#) | [Data](#) | [Publications](#) | [Tutorial](#)

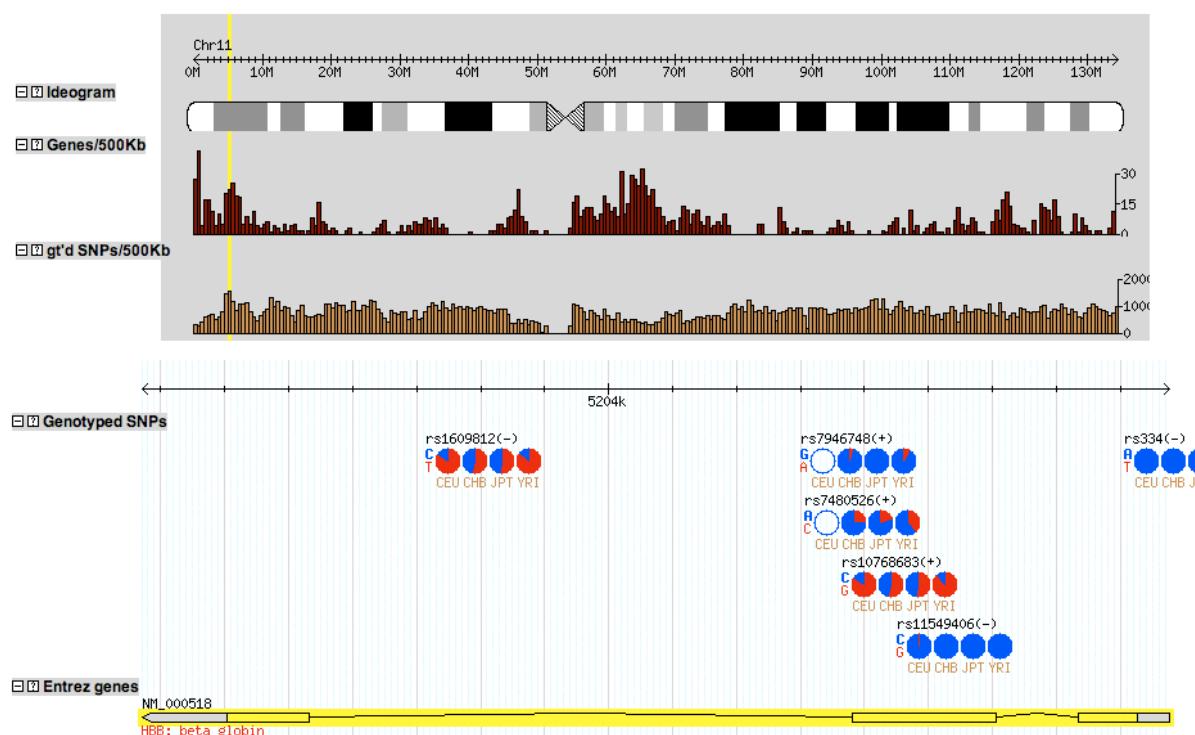
- Inicio en 2002
- Objetivo: construcción de un mapa haplotípico del genoma humano, *HapMap*, que describe los patrones comunes (frecuentes) de variación en la secuencia de DNA humano.
- herramienta clave para buscar genes implicados en patologías humanas y/o en respuestas a fármacos o factores ambientales
- Datos disponibles públicamente

HapMap Fase I (2005):

- 1,3 millones de SNPs (1 SNP / 5 kb)
- Genotipados en 270 individuos de 4 poblaciones:
 - 30 Yorubas (Nigeria)
 - 30 caucasicos (Utha)
 - 45 Han (Beijing)
 - 45 Japoneses

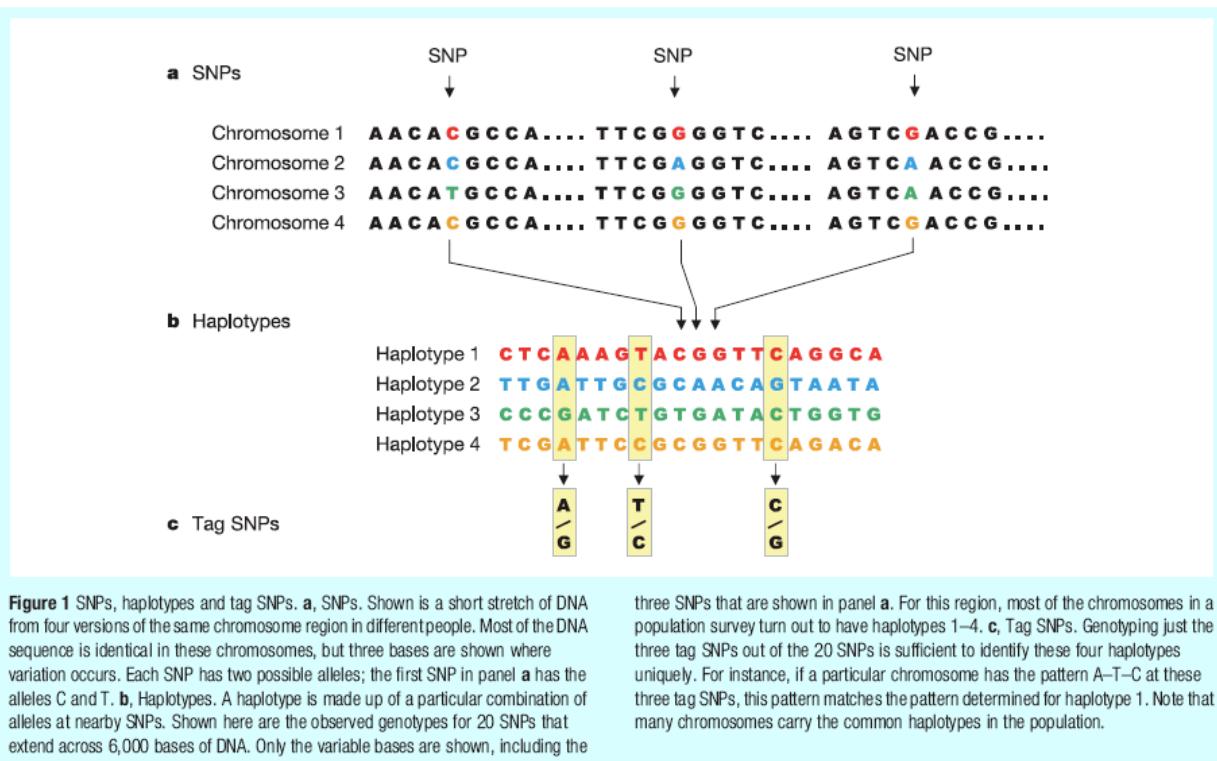
5 febrero 2010

4º curso de genética humana de la SEG



<http://www.hapmap.org/index.html>

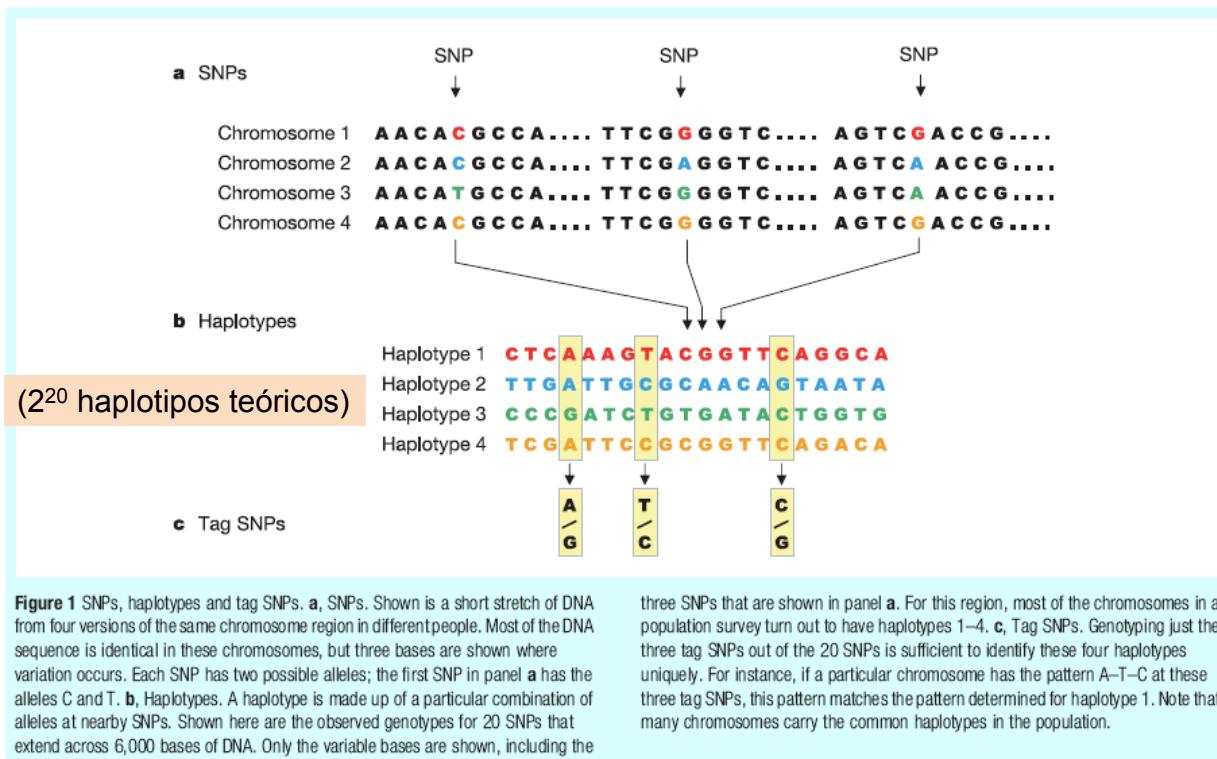
Haplótipos y tag SNPs



Haplótipo:

Combinación concreta de alelos de varios SNPs
presente en un cromosoma dado

Haplótipos

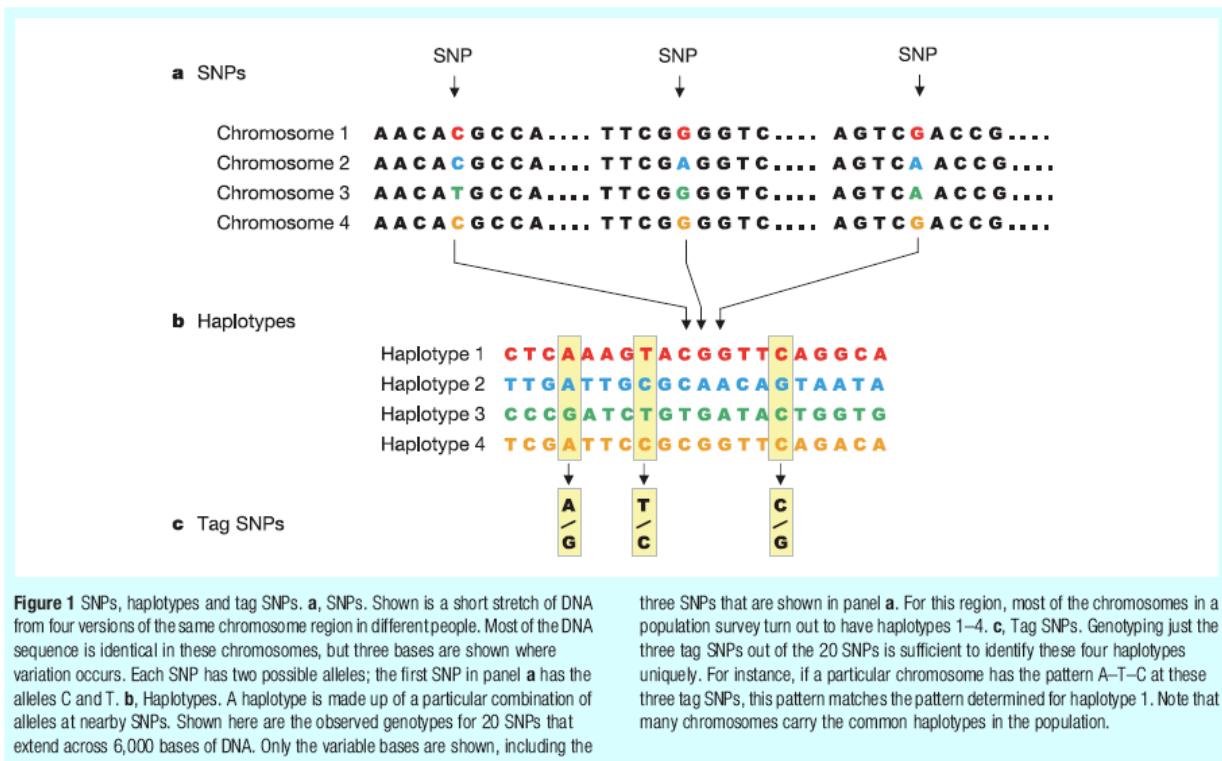


Desequilibrio de Ligamiento

- En una región cromosómica dada se observa un nº pequeño de haplotipos distintos:

Los SNPs muestran una fuerte asociación (*desequilibrio de ligamiento*), dado su origen mutacional y la baja tasa de mutación

tag SNPs



790

© 2003 Nature Publishing Group

NATURE | VOL 426 | 18/25 DECEMBER 2003 | www.nature.com/nature

tagSNPs

- Capturan (casi) toda la variabilidad de un bloque haplotípico sin tener que genotipar todos los SNPs del mismo

HapMap Project

	Phase 1	Phase 2	Phase 3
Samples & POP panels	269 samples (4 panels)	270 samples (4 panels)	1,115 samples (11 panels)
Genotyping centers	HapMap International Consortium	Perlegen	Broad & Sanger
Unique QC+ SNPs	1.1 M	3.8 M (phase I+II)	1.6 M (Affy 6.0 & Illumina 1M)
Reference	Nature (2005) 437:p1299	Nature (2007) 449:p851	Draft Rel. 1 (May 2008)

Phase 3 Samples

label	population sample	# samples	QC+ Draft 1
ASW*	African ancestry in Southwest USA	90	71
CEU*	Utah residents with Northern and Western European ancestry from the CEPH collection	180	162
CHB	Han Chinese in Beijing, China	90	82
CHD	Chinese in Metropolitan Denver, Colorado	100	70
GIH	Gujarati Indians in Houston, Texas	100	83
JPT	Japanese in Tokyo, Japan	91	82
LWK	Luhya in Webuye, Kenya	100	83
MEX*	Mexican ancestry in Los Angeles, California	90	71
MKK*	Maasai in Kinyawa, Kenya	180	171
TSI	Toscans in Italy	100	77
YRI*	Yoruba in Ibadan, Nigeria	180	163
		1,301	1,115

* Population is made of family trios

Variantes Estructurales

- “Cualquier variante que altera la estructura cromosómica ”
- Umbral arbitrario : > 1 kb

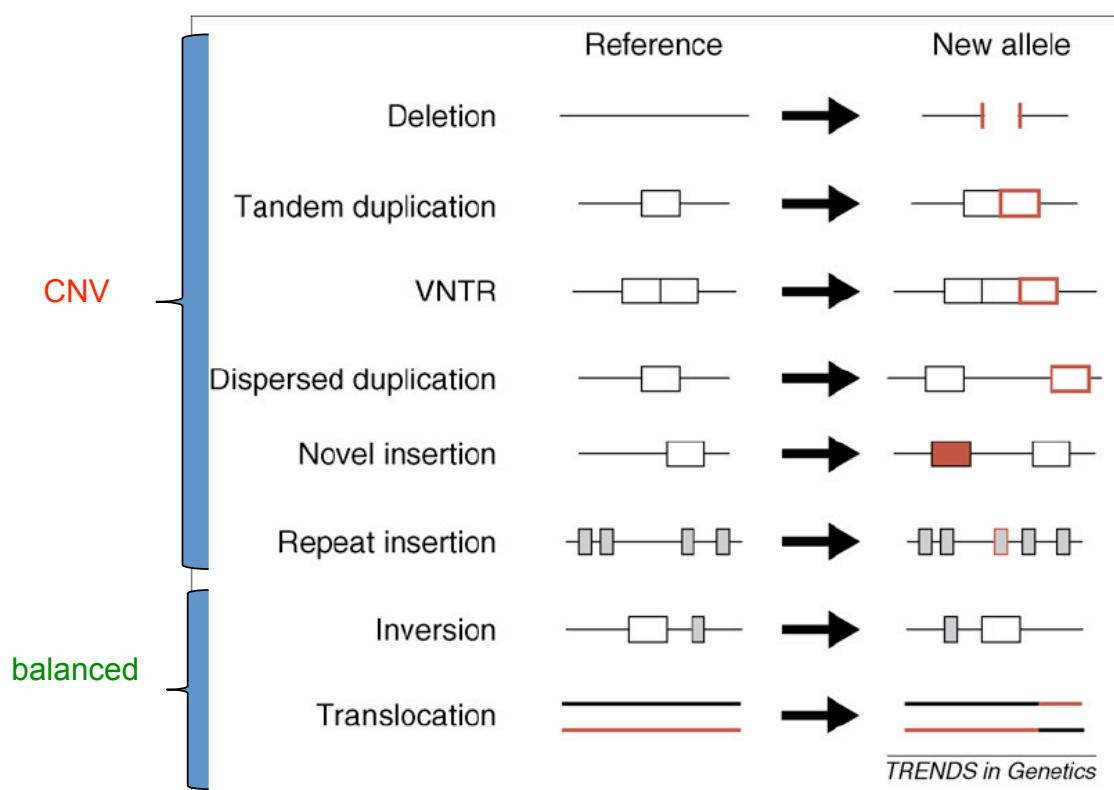


Figure 1. Types of structural variant. Eight different types of structural variant are depicted, defined relative to the reference genome sequence.

CNV (*copy number variants*)

- Abundan en el genoma y están ampliamente distribuidas
- Descubiertas gracias a las nuevas tecnologías genómicas
- Pueden explicar diferencias de expresión génica, enfermedades genéticas y evolución humana

Database of Genomic Variants

<http://projects.tcag.ca/variation/>

Hosted by:
The Centre for
Applied Genomics



[About The Project](#) | [Genome Browser](#) | [Download](#) | [Links](#) | [Data Submissions](#) | [Email us](#)

Please select genome assembly: [Build 36 \(Mar. 2006\)](#)

View Data by Chromosome

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [X](#) [Y](#) [All](#)

View Data by Genome

[Chromosome 1](#) [Chromosome 2](#) [Chromosome 3](#) [Chromosome 4](#) [Chromosome 5](#) [Chromosome 6](#) [Chromosome 7](#) [Chromosome 8](#) [Chromosome 9](#) [Chromosome 10](#) [Chromosome 11](#) [Chromosome 12](#) [Chromosome 13](#) [Chromosome 14](#) [Chromosome 15](#) [Chromosome 16](#) [Chromosome 17](#) [Chromosome 18](#) [Chromosome 19](#) [Chromosome 20](#) [Chromosome 21](#) [Chromosome 22](#) [X](#) [Y](#)

Keyword Search

Exact Match? Yes No
Examples: clone name, accession number, cytoband, gene

BLAT Search

Enter sequence in FASTA format here:

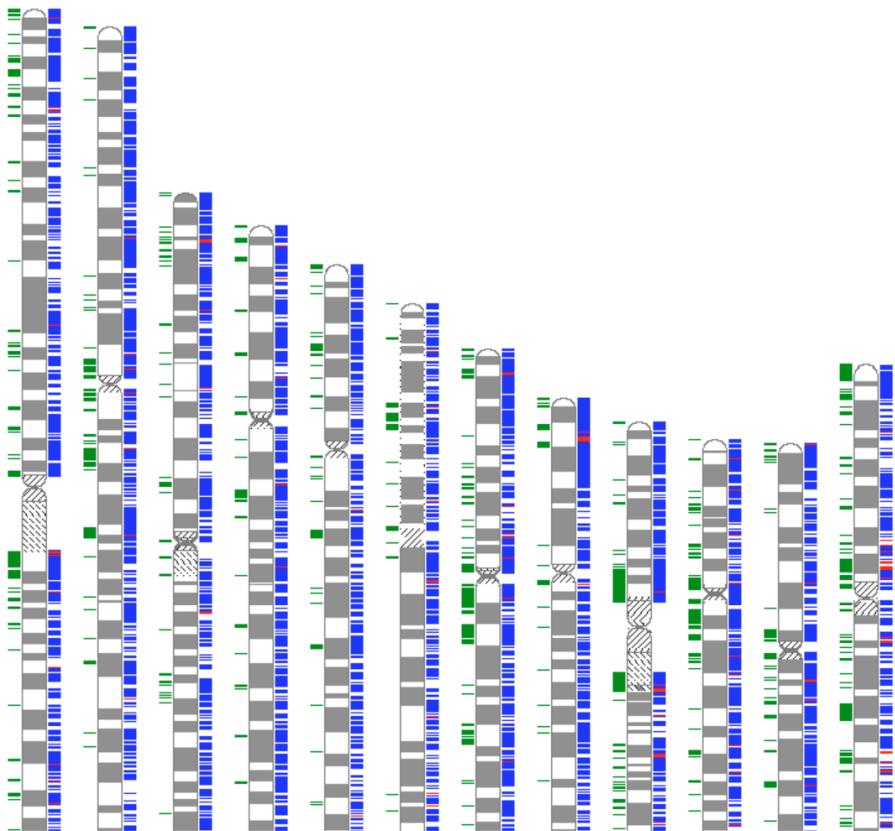
Summary Statistics

Total entries: 49988 (hg18)
CNVs: 29133
Inversions: 914
Indels (100bp-1Kb): 19941
Total CNV loci: 8410
Articles cited: 35
Last updated: Aug 05, 2009
Join our mailing list

[CNV genotype data from the Genome Structural Variation Consortium now available - Aug 21, 2009](#)

Read about the updates in our [newsletter](#).

In citing the Database of Genomic Variants please refer to: Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. *Nat Genet*. 2004 Sep;36(9):949-51.



Nuevas herramientas

ARTICLES

**nature
genetics**

Integrated detection and population-genetic analysis of SNPs and copy number variation

Steven A McCarroll^{1-4,10}, Finny G Kuruvilla^{1-4,10}, Joshua M Korn¹⁻⁶, Simon Cawley⁷, James Nemesh¹, Alec Wysoker¹, Michael H Shapero⁷, Paul I W de Bakker^{1,4,8}, Julian B Maller³, Andrew Kirby³, Amanda L Elliott¹, Melissa Parkin¹, Earl Hubbard⁷, Teresa Webster⁷, Rui Mei⁷, James Veitch⁷, Patrick J Collins⁷, Robert Handsaker¹, Steve Lincoln⁷, Marcia Nizzari¹, John Blume⁷, Keith W Jones⁷, Rich Rava⁷, Mark J Daly^{1,3,4,9}, Stacey B Gabriel¹ & David Altshuler^{1-4,9}

•McCarroll et al. 2001. *Nature genetics* 40:1166-74, 2008 .

Genome-Wide Human SNP Array 6.0



Affymetrix®

- > 906,600 SNPs
- > 946,000 copy number probes
- 474,25 €/muestra

• McCarroll et al. 2001. *Nature genetics* 40:1166-74, 2008 .

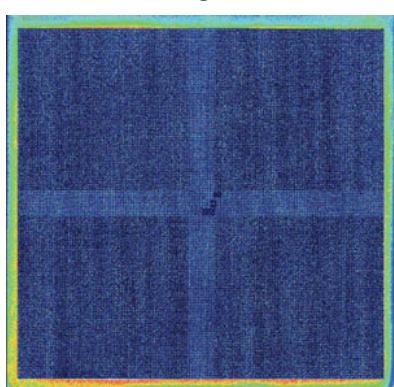
5 febrero 2010

4º curso de genética humana de la SEG

Nuevas herramientas

Nos permiten:

- Abordar estudios de asociación a nivel de genoma completo (GWA)
- Buscar las bases genéticas sin definir hipótesis a priori (*hypothesis free*)
- Mirar simultáneamente SNPs y CNVs
- Estudiar la genética de
 - enfermedades monogénicas
 - enfermedades complejas
 - malformaciones congénitas
 - respuesta a fármacos
 - respuesta a efectos ambientales



5 febrero 2010

4º curso de genética humana de la SEG

Últimas noticias:

Origins and functional impact of copy number variation in the human genome

Donald F. Conrad^{1*}, Dalila Pinto^{2*}, Richard Redon^{1,3}, Lars Feuk^{2,4}, Omer Gokcumen⁵, Yujun Zhang¹, Jan Aerts¹, T. Daniel Andrews¹, Chris Barnes¹, Peter Campbell¹, Tomas Fitzgerald¹, Min Hu¹, Chun Hwa Ihm⁵, Kati Kristiansson¹, Daniel G. MacArthur¹, Jeffrey R. MacDonald², Ifejenelo Onyiah¹, Andy Wing Chun Pang², Sam Robson¹, Kathy Stirrups¹, Armand Valsesia¹, Klaudia Walter¹, John Wei², Wellcome Trust Case Control Consortium†, Chris Tyler-Smith¹, Nigel P. Carter¹, Charles Lee⁵, Stephen W. Scherer^{2,6} & Matthew E. Hurles¹

Structural variations of DNA greater than 1 kilobase in size account for most bases that vary among human genomes, but are still relatively under-ascertained. Here we use tiling oligonucleotide microarrays, comprising 42 million probes, to generate a comprehensive map of 11,700 copy number variations (CNVs) greater than 443 base pairs, of which most (8,599) have been validated independently. For 4,978 of these CNVs, we generated reference genotypes from 450 individuals of European, African or East Asian ancestry. The predominant mutational mechanisms differ among CNV size classes. Retrotransposition has duplicated and inserted some coding and non-coding DNA segments randomly around the genome. Furthermore, by correlation with known trait-associated single nucleotide polymorphisms (SNPs), we identified 30 loci with CNVs that are candidates for influencing disease susceptibility. Despite this, having assessed the completeness of our map and the patterns of linkage disequilibrium between CNVs and SNPs, we conclude that, for complex traits, the heritability void left by genome-wide association studies will not be accounted for by common CNVs.